LECTURE 1

- POPULAR SCIENCE AND ALGORITHMS. → AND HOW THEY ARE DIFFERENT FROM COMP 414/514. (DISCLAIMER: WE WILL NOT COVER THEM IN DEPTH)

- MULTI-AGENT LEARNING ALGORITHM: HOW A COMPUTER CAN PLAY STARCRAFT?

KEYWORDS: POLICY GRADIENT METHODS

E.G.: $\underset{\theta}{\text{maximize}}\ \hat{\mathbb{E}}_t\left[\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}\cdot\hat{A}_t\right]$

π_θ IS STOCHASTIC POLICY
$\hat{A}_t$: ESTIMATOR OF ADV. FUNCTION
$\hat{\mathbb{E}}_t$: EMPIRICAL AVERAGE

S.T. $\hat{\mathbb{E}}_t\left[KL\left[\pi_{\theta_{old}}(\cdot|s_t), \pi_\theta(\cdot|s_t)\right]\right] \le \delta$ ← TRPO

SOURCE: "PROXIMAL POLICY OPTIMIZATION ALGORITHMS".

OPENAI-FIVE: SCALED VERSION OF PROXIMAL POLICY OPTIMIZATION
+
LSTMS FOR EACH HERO
=
DEFEAT AMATEURS AT DOTA 2. (VIDEO)

] OF 180 YEARS OF PLAY WITHIN A DAY.

ALPHASTAR: BLIZZARD + DEEPMIND FOR STARCRAFT

ALSO CHECK: HOW TO PLAY TEXAS HOLD'EM, ...

(CONSTRAINED OPTIMIZATION / BEYOND SIMPLE SGD)

- GENERATIVE ADVERSARIAL LEARNING

KEYWORDS: IMAGE SYNTHESIS USING GANS

SOURCE: "SEMANTIC IMAGE SYNTHESIS WITH SPATIALLY-ADAPTIVE NORMALIZATION"

GauGau: SYNTHESIZE IMAGES FROM SEGMENTATION MAPS (VIDEO)

(MIN-MAX PROBLEMS / BEYOND SGD)

- NEURAL MACHINE TRANSLATION

KEYWORDS: CROSS-LINGUAL PRETRAINING, SELF-SUPERVISED LEARNING. BACK-TRANSLATION, TRANSFORMERS,

SOURCE: "FB FAIR'S WMT19 NEWS TRANSLATION TASK SUBMISSION"

- FAST COMBINATORIAL OPTIMIZATION ALGORITHS

KEYWORDS: (INSPIRED BY) QUANTUM ADIABATIC OPTIMIZATION, BIFURCATION, ISING MODELS (= MAXCUT).

(DISCRETE OPTIMIZATION)

SOURCE : "COMBINATORIAL OPTIMIZATION BY SIMULATING ADIABATIC BIFURCATIONS IN NONLINEAR HAMILTONIAN SYSTEMS"

SB ALGORITHM : - 2.000 NODES ⟶ 0.5 SEC MAXCUT SOLUTION (FPGA)

- 100K NODES ⟶ FEW SECS -11- (GPU CLUSTER)

(FULLY CONNECTED GRAPHS.)

- QUANTUM COMPUTING & QUANTUM ALGORITHMS

(A DIFFERENT COMPUTATIONAL MODEL)

KEYWORDS: QUANTUM SUPREMACY, QUANTUM APPLICATIONS. (FACTORIZATION, SEARCH, TELEPORTATION, ...),

SOURCE : (MULTIPLE)

- "QUANTUM SUPREMACY USING A PROGRAMMABLE SUPERCONDUCTING PROCESSOR"
- "ON "QUANTUM SUPREMACY" " (IBM BLOG).

SYCAMORE : GOOGLE'S QUANTUM COMPUTER

53 QBUTIS

PROBLEM : FIGURE OUT PROBABILITY DISTRIBUTION OF ALL POSSIBLE OUTCOMES FROM A QUANTUM RANDOM NUMBER GENERATOR.

(GOOGLE CLAIMED TO HAVE USED THIS SEEMINGLY USELESS APPLICATION IN PRACTICE).

SUPREMACY : IBM'S LARGEST HPC CONFIGURATION (SUMMIT.)

TAKES 10,000 YEARS ⟶ (IBM) 2,5 DAYS

SUMMIT VS SYCAMORE SIZE : A ROOM VS. 2 BASKETBALL COURTS.

PROBLEMS IN QUANTUM: PURE QUANTUM (QUANTUM INFORMATION / REPRESENTATION MANIPULATION) SHOR'S, GROVER.... VS HYBRID (QAOA)

$\langle \cdot | U | \cdot \rangle$ s.t. $|U\rangle$ IS A QUANTUM GATE

- OVERVIEW OF COMP 414/514.
  - GRADIENT DESCENT : CONVEX & NON-CONVEX.
  - CONDITIONAL GRADIENT (FRANK WOLFE)
  - NEWTON'S METHOD, QUASI-NEWTON TECHNIQUES
  - MOMENTUM METHODS
  - SGD & VARIANTS.
  - SPARSITY-BASED METHODS
  - LOW-RANK METHODS

  → NON-CONVEX ALGORITHMS

  - LANDSCAPE OF SMOOTH FUNCTIONS
  - DISTRIBUTED IMPLEMENTATIONS

COMMON THEME: THEORY + PRACTICE.

OVERVIEW OF COMP 545.

- ALGORITHMS FOR **MODERN SMOOTH UNCONSTRAINED OPTIMIZATION** (ADAGRAD, RMSPROP, ADAM)
- CONSTRAINED OPTIMIZATION & DUALITY THEORY
- MIRROR DESCENT & MWU ALGORITHM.
- INTERIOR POINT METHODS (PRIMAL MOSTLY)
- FOCUS: ADVERSARIAL ROBUSTNESS AND STABILITY.
- SADDLE-POINT OPTIMIZATION, MIN-MAX, GANS.
- DISCRETE OPTIMIZATION TO CONVEXITY, AND BACK (SPECIAL TALK: MAXCUT AND HYPERSPHERICAL COORDINATES)
- QUANTUM COMPUTATION: INTRO, FAMOUS ALGORITHMS, HYBRID ALGO'S
- UNCERTAINTY IN ML: INFERENCE, GPs.

(BACKUP TOPICS: *- SUPERVISED LEARNING, DFO METHODS, SKETCHING + LINEAR ALGEBRA, LATEST NN ARCHITECTURES).

- LOGISTICS.
  - IN CONTRAST TO COMP 414/514: **NO SLIDES, NO CODE** - ONLY HANDWRITTEN NOTES
  - FOR GRADUATE STUDENTS MOSTLY.

- FORMAT: LECTURE BY ME ON A TOPIC, FOLLOWED BY PRESENTATION SESSION (ASK OPINION ON SEMINAR COURSES).
  (GAME: I WILL BE THE ADVERSARY)
- WORKLOAD: 1 CREDIT → 1 PRESENTATION (100%)
  3 CREDIT → FINAL PROJECT (80% - 20%) → LITERATURE REVIEW / RESEARCH
  (I MIGHT ADD SCRIBING).
  REVIEWS?
- GOALS: LEARN ONLY A BIT ABOUT ALGORITHMS IN RECENT RESEARCH.
- PREREQ: CALCULUS, LIN. ALGEBRA, ML
  COMP 414/514 WOULD BE IDEAL → FUTURE GOAL: CREATE CHAINS OF COURSES.
  QUIZ WILL BE GIVEN.
- CONTACT INFO: ricecomp414@gmail.com.
  NO CANVAS

Quiz (15 min.)

- RECOMMENDATION: READ NOTES FROM COMP 414/514.

LECTURES 2

- ALGORITHMS IN NN TRAINING
  * READING ASSIGNMENTS: DECAYING MOMENUM
    AMSGRAD.
    ...
    ADAPTIVE METHODS - MARGINAL...
    MINIMUM NORM SOLUTIONS ...

OVERVIEW OF COMP 414/514

- GD, CONVERGENCE RATES IN $O(\cdot)$, SOME PROPERTIES OF FUNCTIONS, CONVEXITY SECOND-ORDER METHODS, MOMENTUM, (SGD, COORDINATE DESCENT, FRANK-WOLFE), LANDSCAPE OF FUNCTIONS,

INTERLUDE

- GENERAL OPTIMIZATION

$$\min_{x \in \mathbb{R}^p} f(x) \qquad \text{OR} \qquad \min_{x \in \mathbb{R}^p} f(x) \quad \text{s.t.} \quad g(x) \leq 0$$

PROVIDE SOME EXAMPLES:

1. LEAST SQUARES.
2. MATRIX REGRESSION
3. 2-HIDDEN LAYER MLP

YOU SHOULD KNOW: VECTOR NOTATION + PROPERTIES
INNER PRODUCT
NORMS (l_2, l_1, l_∞, l_0)
MATRIX NOTATION
MATRIX MULTIPLICATION
INNER PRODUCT (MATRIX)
EIG / SVD

- GRADIENT DESCENT.

YOU SHOULD KNOW: DERIVATIVES & GRADIENT
HESSIAN AND ITS GEOMETRICAL INTERPRETATION
$\nabla^2 f(\cdot) \succ 0, \prec 0, \succcurlyeq 0 \ldots$
TAYLOR APPROXIMATION
LIPSCHITZ CONDITION
-//- GRADIENT CONDITION

GD: $x_{t+1} = x_t - \eta_t \nabla f(x_t)$ ⟶ WHAT CAN WE SAY ABOUT:
i) η_t?
ii) INITIALIZATION?
iii) CONVERGENCE?

CONVERGENCE THM: CLAIM 6.

OVERVIEW OF CONV. RATES.

- CONVEXITY.

DEFINITION: $f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y)$, $\alpha \in [0,1]$

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle$$

GRADIENT DESCENT UNDER CONVEXITY: COMPARE WITH ABOVE

STRONG CONVEXITY AND GUARANTEES.

LOWER BOUNDS

- CONVEX SETS & PROJECTIONS

- BEYOND FIRST-ORDER METHODS
  - NEWTON'S METHOD. + GUARANTEES
  - QUASI-NEWTON METHODS

- MOMENTUM ACCELERATION.

HEAVY BALL METHOD VS NESTEROV SCHEMAS

GUARANTEES

SGD.